Mathematical Modeling and Optimization of Cellulase Protein Production Using *Trichoderma reesei* RL-P37

Arun Tholudur,[1] W. Fred Ramirez,[1] James D. McMillan[2]

[1]Department of Chemical Engineering, University of Colorado, Boulder, Colorado 80309, USA; telephone: (303)-492-8660; fax: (303)-492-4341; e-mail: fred.ramirez@colorado.edu
[2]Biotechnology Center for Fuels and Chemicals, National Renewable Energy Laboratory, Golden, Colorado, USA

Received 4 April 1999; accepted 30 June 1999

Abstract: The enzyme cellulase, a multienzyme complex made up of several proteins, catalyzes the conversion of cellulose to glucose in an enzymatic hydrolysis-based biomass-to-ethanol process. Production of cellulase enzyme proteins in large quantities using the fungus *Trichoderma reesei* requires understanding the dynamics of growth and enzyme production. The method of neural network parameter function modeling, which combines the approximation capabilities of neural networks with fundamental process knowledge, is utilized to develop a mathematical model of this dynamic system. In addition, kinetic models are also developed. Laboratory data from bench-scale fermentations involving growth and protein production by *T. reesei* on lactose and xylose are used to estimate the parameters in these models. The relative performances of the various models and the results of optimizing these models on two different performance measures are presented. An approximately 33% lower root-mean-squared error (RMSE) in protein predictions and about 40% lower total RMSE is obtained with the neural network-based model as opposed to kinetic models. Using the neural network-based model, the RMSE in predicting optimal conditions for two performance indices, is about 67% and 40% lower, respectively, when compared with the kinetic models. Thus, both model predictions and optimization results from the neural network-based model are found to be closer to the experimental data than the kinetic models developed in this work. It is shown that the neural network parameter function modeling method can be useful as a "macromodeling" technique to rapidly develop dynamic models of a process. © 1999 John Wiley & Sons, Inc. *Biotechnol Bioeng* **66:** 1–16, 1999.
Keywords: kinetic modeling; neural networks; parameter function modeling; cellulase production; *Trichoderma reesei*

Correspondence to: W. F. Ramirez
Contract grant sponsors: U.S. Department of Energy; NREL; The Whitaker Foundation
Contract grant number: KAK-6-16810-06

INTRODUCTION

Enzymatic hydrolysis of cellulose to glucose is carried out by the enzyme cellulase, a multienzyme complex made up of several proteins. The fungus *Trichoderma reesei* is an efficient producer of cellulase enzymes. Production of cellulase enzyme in large quantities requires understanding and proper controlling of the growth and enzyme production capabilities of *T. reesei*. This is an extremely complicated system; many factors influence the organism's ability to grow and produce enzyme.

Strains of *T. reesei* have been studied in great detail for more than two decades for their ability to produce the cellulase enzyme complex (Esterbauer et al., 1991; Kadam, 1996; Philippidis, 1994). Most literature reports describe experiments that have achieved a significant enzyme titer or a high volumetric productivity. Only a few publications deal with modeling cellulase production systems.

Most publications provide performance data for specific strains under particular conditions without any attempt to model the system. For example, Mohagheghi et al. (1988) carried out a series of experiments that examined varying the amounts of xylose and cellulose in batch cultures of *T. reesei* mutant RUT-C30. They also performed a series of fed-batch experiments (Mohagheghi et al., 1990) using the same medium constituents. Chaudhuri and Sahai (1993) examined enzyme production using *T. reesei* C5 growing on lactose in batch culture. Tangnu et al. (1981) studied the effects of various process parameters, such as temperature, pH, carbon sources, and substrate concentration on the ratio of mycelial growth and extracellulose enzyme production using *T. reesei* RUT-C30. Schafner and Toledo (1992) investigated cellulase enzyme production in continuous culture by *T. reesei* strain QM9414 on a xylose-based medium. They also studied enzyme production on a xylose-based medium supplemented with sorbose (Schafner and Toledo, 1991). Lejeune and Baron (1995) assessed the effect of agitation on growth and enzyme production by *T. reesei*

 CCC 0006-3592/99/010001-16

QM9414. Hendy et al. (1984) reported using fed-batch cultivation to enhance cellulase enzyme production using the RUT-C30 strain. Also, many reports have demonstrated enzyme production on cellulosic biomass rather than mineral-based media supplemented with pure carbon sources. To name a few, Acebal et al. (1988) used *T. reesei* QM9414 to produce cellulase from wheat straw, whereas Szengyel et al. (1997) studied cellulase production using hemicellulose hydrolyzate from steam-pretreated willow.

Modeling the growth and product formation characteristics of various microorganisms is a very challenging task. There are many different approaches to modeling the microbial kinetics (for an excellent review, see Nielsen and Villadsen, 1992). Simplistic unstructured models (e.g., incorporating Monod kinetic terms or other empirical expressions) do not offer much in terms of elucidating the exact nature of these processes. However, more structured models often involve introducing process variables that cannot be estimated reliably. Several studies dealing with modeling the growth and morphology of complex microorganisms such as filamentous fungi have been published in recent years (Lejeune and Baron, 1998; Nielsen, 1992, 1996).

Among the publications that consider the development of models for cellulase enzyme production, probably the most comprehensive model was presented by Bader et al. (1993), who modeled the growth and enzyme production characteristics of *T. reesei* RUT-C30 using potato pulp as the cellulosic substrate. Their model accounted for physiological changes in the fungal cells that better described the enzyme production behavior. The model also incorporated certain key features such as adsorption of enzyme onto the substrate leading to its decomposition and hydrolysis, followed by cell growth due to the uptake of the resulting simple sugars. Another important feature in this model was the classification of the substrate into amorphous and crystalline cellulosic components, a hemicellulosic component, and other nondegradable components.

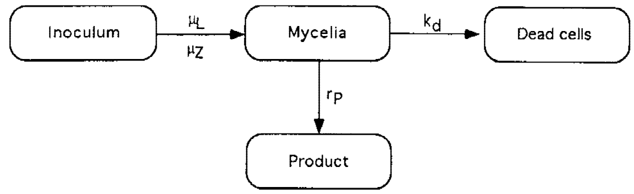


Figure 1. Schematic representation of the unstructured kinetic model.

Rakshit and Sahai (1991) formulated an empirical model for growth of and enzyme production by the *T. reesei* mutant E-12 and developed an optimal control strategy for enhancing production of the enzyme. However, they chose not to model substrate consumption and hence their model did not consider the dependence of cell growth on the substrate concentration. Chaudhuri and Sahai (1994) estimated some growth and maintenance parameters for cellulase biosynthesis on *T. reesei* C5 and compared them with some published data. Using data from continuous culture experiments, they demonstrated that Monod's equation adequately represented the growth dynamics.

More recently, Velkovska et al. (1997) developed a kinetic model for batch cellulase production by RUT-C30 using solka floc (purified cellulose) as the substrate. Their approach resulted in a model that was simpler than that of Bader et al. (1993) by ignoring the actual composition of the enzyme complex. Their model took into account the morphological and physiological changes that occur during growth of *T reesei* cells, and was structured in biomass and substrate reactivity variables but unstructured with respect to intracellular mechanisms and molecular species.

Although many data are available for strains of *T. reesei*, such as RUT-C30, not much work has been done in quantifying the dynamics of other strains such as RL-P37. In this work, we develop three models for growth and enzyme production by *T. reesei* RL-P37 on a soluble sugar substrate system (lactose and xylose). The first two are kinetic models incorporating different degrees of structure. The third is based on the neural network parameter function approach. The motivation for using neural network parameter functions is to improve the accuracy and predictive capabilities of kinetic models (Tholudur and Ramirez, 1996), and this approach has been successfully applied to protein production by recombinant bacteria (Tholudur and

Table I. Growth media composition.

Component	Stage 1	Stage 2	Fermentor
Salts			
KH_2PO_4 (g/L)	3.8	3.8	3.8
$MgSO_4 \cdot 7H_2O$ (g/L)	0.6	0.6	0.6
$CaCl_2 \cdot 2H_2O$ (g/L)	0.8	0.8	0.8
$(NH_4)_2SO_4$ (g/L)	5	5	5, 7.5, 10, 12.5, 15
Trace minerals			
$FeSO_4 \cdot 7H_2O$ (mg/L)	5	5	5
$MnSO_4 \cdot H_2O$ (mg/L)	1.6	1.6	1.6
$ZnSO_4 \cdot 7H_2O$ (mg/L)	1.4	1.4	1.4
$CoCl_2 \cdot 6H_2O$ (mg/L)	3.7	3.7	3.7
Protein supplementation			
Corn steep liquor (% v/v)	2	2	2
Carbon source			
Glucose (% w/v)	2	—	—
Xylose (% w/v)	—	1	1
Lactose (% w/v)	—	1	1, 2, 3, 4, 5
Miscellaneous			
Tween-80 (mL/L)	—	—	0.2
Antifoam (mL/L)	—	—	0.1

Table II. Estimated parameters for the unstructured kinetic model.

Parameter	Value	Parameter	Value
μ_{max},L	0.0493 h^{-1}	Y_L	0.3770 g/g
$K_{S,L}$	0.2346 g/L	Y_Z	0.6842 g/g
$K_{I,L}$	50.26 g/L	k_d	0.0065 h^{-1}
μ_{max},Z	0.1253 h^{-1}	α	0.0409
$K_{S,Z}$	5.734 g/L	β	0.0015 h^{-1}

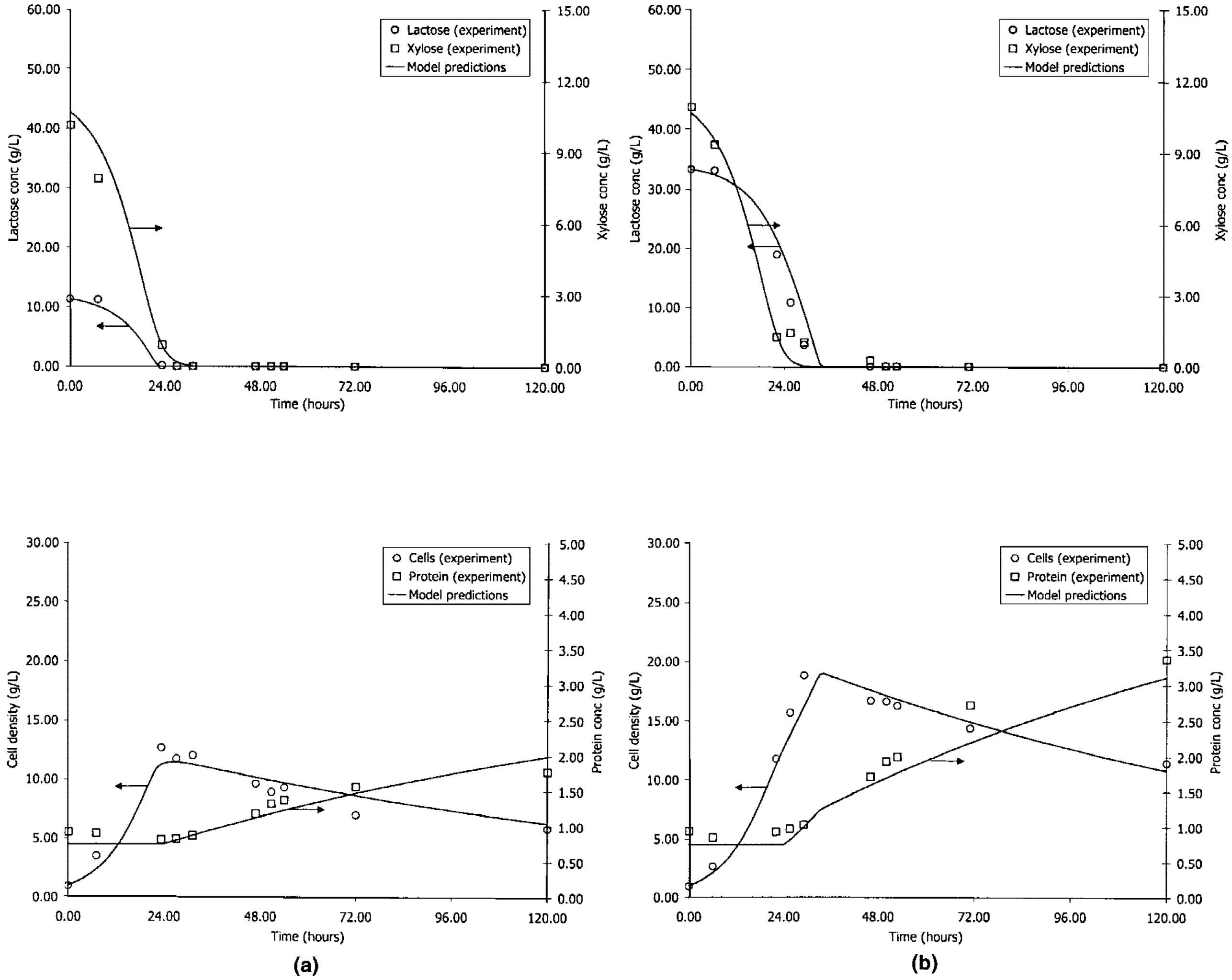


Figure 2. (a) Growth and production kinetics as predicted by the unstructured kinetic model (initial xylose = 1% w/v, initial lactose = 1% w/v). (b) Growth and production kinetics as predicted by the unstructured kinetic model (initial xylose = 1% w/v, initial lactose = 3% w/v). (c) Growth and production kinetics as predicted by the unstructured kinetic model (initial xylose = 1% w/v, initial lactose = 5% w/v).

Ramirez, 1999). Here, we report on extending the application of this approach to cellulase production by *T. reesei* RL-P37.

Because the neural network parameter function modeling method can be considered a "macromodeling" approach that utilizes just fundamental macroscopic material balances (i.e., no information on the structure of the system at the microscopic level is incoporated), it is reasonable to attempt the development of kinetic models of similar complexity for comparison of the different approaches. Furthermore, the incorporation of fundamental process knowledge in the neural network parameter function modeling method makes it a hybrid modeling scheme and moves it out of the realm of pure "black-box" modeling and puts it on par with the other kinetic modeling methods discussed in this article.

The following sections describe the experiments conducted and the development of the three mathematical models, followed by a comparison of the models and their application to predict optimal operating conditions.

MATERIALS AND METHODS

Microorganism

Trichoderma reesei RL-P37 (obtained from B.S. Montenecourt, Lehigh University) was used in this work and was maintained on V8-based media plates as suggested by Schell et al. (1990).

Culture Medium and Cultivation

A two-stage preinoculum culture growth method was used. Stage 1 consisted of a 500-mL baffled shake flask with a 100-mL working volume, and stage 2 consisted of 2000-mL baffled shake flasks with a 350-mL working volume. One milliliter of frozen stock culture was transferred to stage 1 as preinoculum, and 17.5 mL of the growth culture was then transferred to stage 2 as inoculum for the fermentor. Table I summarizes the media constituents for the various stages.

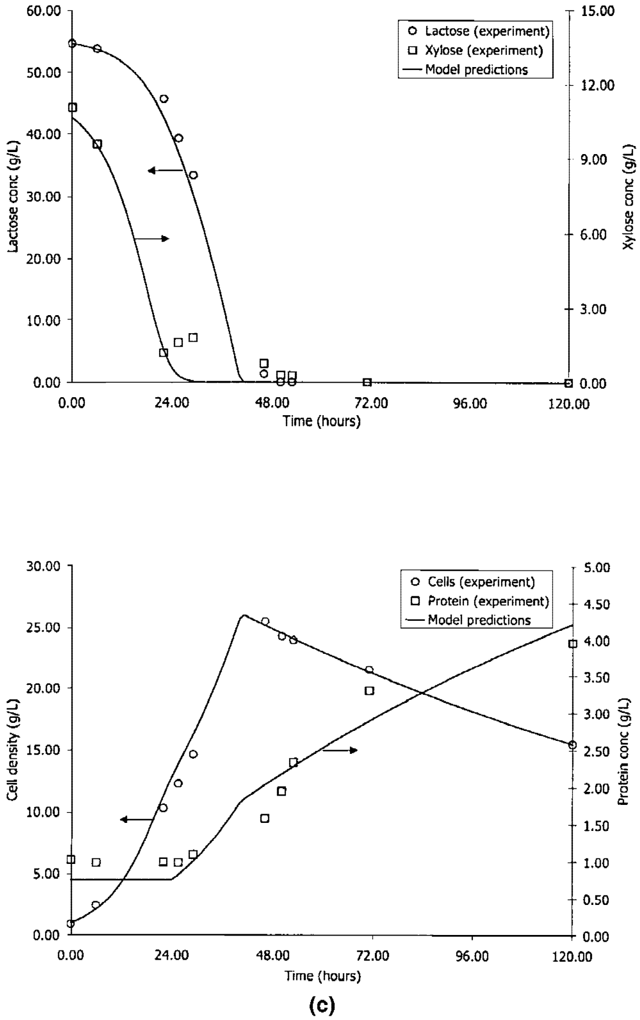


Figure 2. Continued.

New Brunswick Scientific BioFlo 3000 fermentors equipped with temperature, pH, and dissolved oxygen (DO) control were used. Fermentations were run at a working volume of 3 L in 5-L vessels. Three hundred milliliters of the stage 2 culture was used to inoculate the fermentors. The temperature was kept constant at 28°C, agitation was maintained at 450 rpm, and the pH was maintained at 4.8 using concentrated NH_4OH and H_3PO_4. The level of DO was maintained at approximately 20% of air saturation using air supplemented with pure oxygen, except for periods of rapid growth when DO levels fell below 20%.

T. reesei can grow on a variety of substrates. However, most strains of this fungus do not produce the cellulase enzyme complex in appreciable quantities unless they are grown on the insoluble substrate cellulose. The use of insoluble substrate leads to difficulties with accurate determination of cell mass in the presence of other insoluble components. Thus, keeping in mind that the aim of this work was to demonstrate the applicability of the neural network parameter function modeling technique, a soluble sugar substrate system was employed. Lactose, a soluble sugar, also induces the production of cellulase, although to a lesser

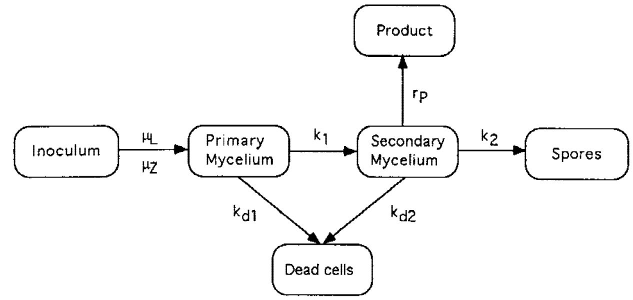


Figure 3. Schematic representation of the structured kinetic model.

extent than cellulose. To add complexity to the modeling endeavor, xylose was also used, and the models were required to capture the preferential uptake of xylose over lactose.

The experimental design consisted of varying the initial levels of lactose with a constant initial xylose concentration. The initial xylose concentration was maintained at approximately 1% w/v and the initial lactose concentration was varied. Five fermentors with initial lactose levels of approximately 1%, 2%, 3%, 4%, and 5% w/v were run. As shown in Table I, the initial amounts of $(NH_4)_2SO_4$ were varied so as to obtain an initial C:N ratio close to 8 (grams C:grams N).

One of the experiments (3% w/v initial lactose concentration) was repeated and it was found that the growth and protein production data reproduced fairly well. Hence, it was deemed unnecessary to repeat the rest of the lactose concentrations.

Analytical Techniques

All fermentations were run for 120 h. Samples of about 40 mL were taken frequently, and typically ten samples were obtained for each fermentation over the course of the experiment. Samples were analyzed for dry cell weight, soluble sugar concentrations, and protein levels, as explained in what follows.

Cell Mass

Cell concentrations (as grams dry cell mass per liter [g DCM/L]) were determined using preweighed Gelman glass

Table III. Estimated parameters for the structured kinetic model.

Parameter	Value	Parameter	Value
$\mu_{max,L}$	0.1008 h^{-1}	k_1	0.0610 h^{-1}
$K_{S,L}$	0.0194 g/L	k_2	0.0069 h^{-1}
$K_{I,L}$	6.502 g/L	k_{d1}	0.0099 h^{-1}
Y_L	0.4106 g/g	k_{d2}	0.0089 h^{-1}
$\mu_{max,Z}$	0.2016 h^{-1}	α	0.0525
$K_{S,Z}$	4.993 g/L	β	0.0025 h^{-1}
Y_Z	0.6418 g/g		

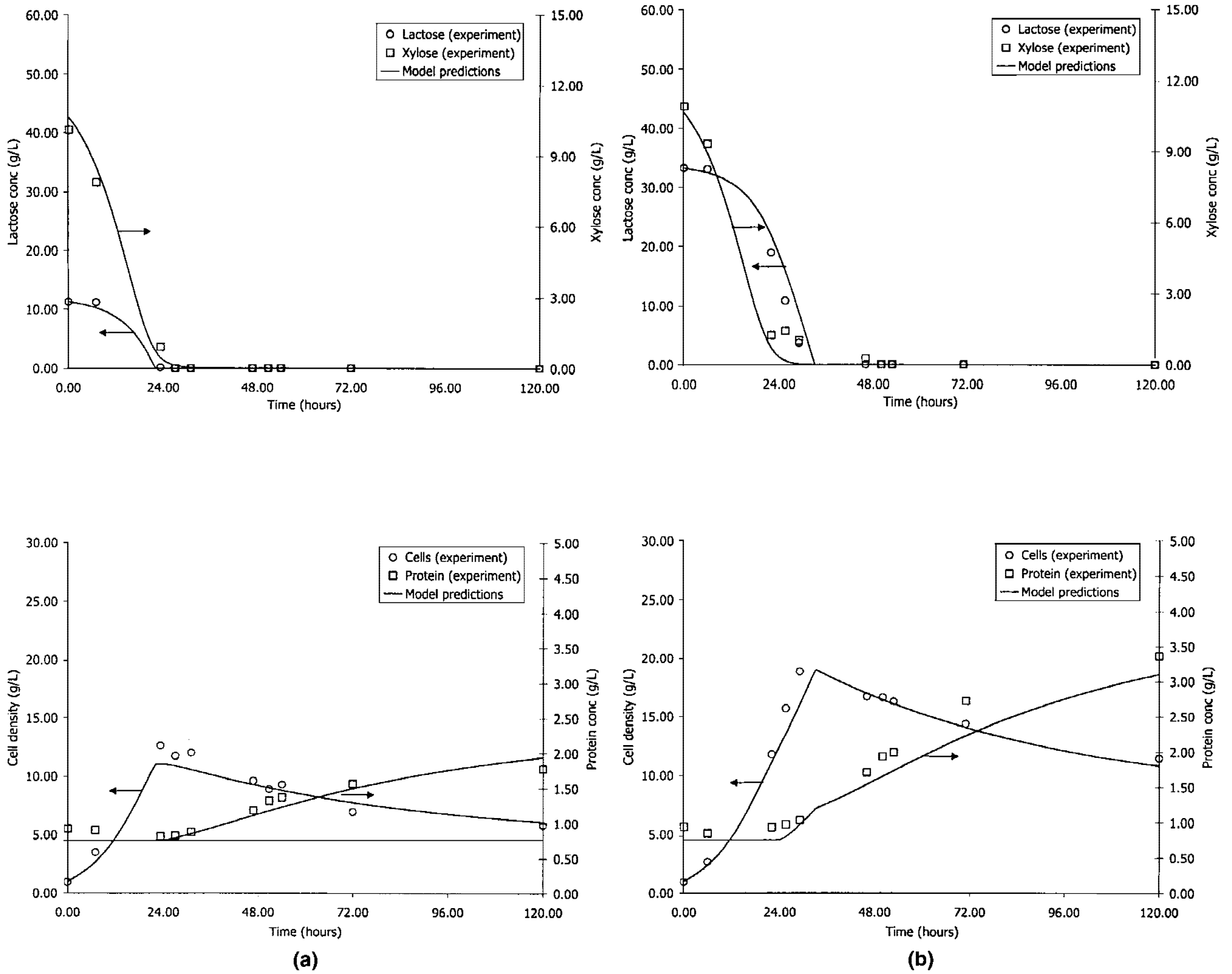


Figure 4. (a) Growth and production kinetics as predicted by the structured kinetic model (initial xylose = 1% w/v, initial lactose = 1% w/v). (b) Growth and production kinetics as predicted by the structured kinetic model (initial xylose = 1% w/v, initial lactose = 3% w/v). (c) Growth and production kinetics as predicted by the structured kinetic model (initial xylose = 1% w/v, initial lactose = 5% w/v).

fiber filters. Four milliliters of the fungal samples were filtered onto the membrane, washed twice with deionized water, and allowed to dry at room temperature under vacuum. Triplicates of the analysis were performed for each sample to estimate the reproductability of the cell concentration measurement.

Substrate

Sugar concentrations (lactose and xylose) were determined using an HPLC (HP 1090 with CHEMSTATION software) equipped with a Bio-Rad organic acids column (HPX-87H, 300 × 7.8 mm) operating at 65°C. Dilute sulfuric acid (0.01N) was used as the mobile phase at a flow rate of 0.6 mL/min.

Protein

Protein concentrations in the culture supernatants were estimated by UV absorbance following chromatographic fractionation using a 5-mL Pharmacia Hi-Trap column packed with Sephadex G-25. This is a modification of the method described by Adney et al. (1995). A Beckman System Gold HPLC system equipped with the programmable Model 126 solvent module, Model 507 autosampler, and a Model 166 variable wavelength detector was used to separate proteins from media components and low-molecular-weight proteins. The Beckman System Gold software was used to integrate peak areas and determine peak heights.

Samples were diluted and subjected to chromatography in 20 m*M* acetate, 100 m*M* NaCl (pH 5.0) buffer. The system's autosampler was then used to inject the samples in 100-μL volumes with the flow rate maintained at 2 mL/min. Sephadex G-25 has an exclusion limit of approximately 5000 Da and can be used for the group separation of proteins >5000 Da from smaller peptides typically present in the nutrient media. Only the void volume peak was assumed to be protein >5000 Da and was included in the total protein calculation calibrated using bovine serum albumin standard so-

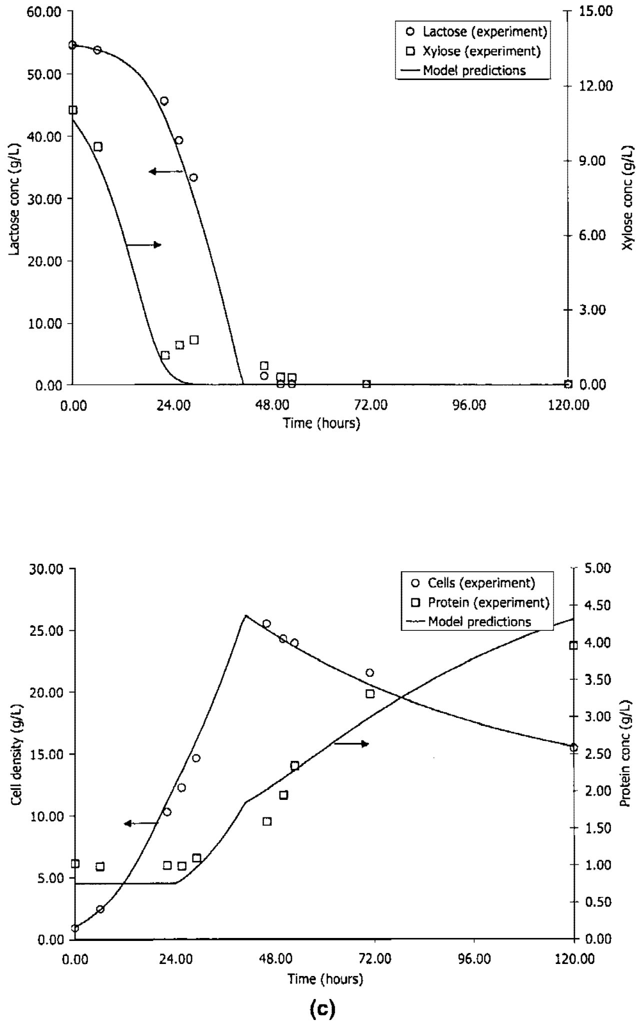


Figure 4. Continued.

lutions. This method provides a reasonable estimation of the total hydrolytic enzyme concentration based on the assumption that only proteins produced by the fungus and secreted into the culture broth, and not proteins present in the nutrient medium, should contribute to the absorbance in the void volume of the column.

MODEL DEVELOPMENT

Three models of the growth and protein production dynamics are developed in this section.

Unstructured Kinetic Model

Model Development

The simplest kinetic model assumes that sugars are converted into mycelial cell mass, which produces enzyme as shown in Figure 1. The differential equations describing this system are as follows:

$$\frac{dL}{dt} = -\mu_L X / Y_L \tag{1}$$

$$\frac{dZ}{dt} = -\mu_Z X / Y_Z \tag{2}$$

$$\frac{dX}{dt} = \mu_L X + \mu_Z X - k_d X \tag{3}$$

$$\frac{dP}{dt} = r_P X \tag{4}$$

where L, Z, X, and P are the lactose, xylose, cell mass, and protein concentrations, respectively. The specific growth rates on lactose and xylose are μ_L and μ_Z, respectively; Y_L and Y_Z are the corresponding cell mass yields on lactose and xylose. k_d is an endogenous death term and r_P is the specific protein production rate. When both lactose and xylose are present in the medium, xylose is preferentially taken up. The following forms are postulated for the specific growth rate and protein production rate:

$$\mu_L = \frac{\mu_{\max,L} L}{K_{S,L} + L} \frac{K_{I,L}}{K_{I,L} + Z} \tag{5}$$

$$\mu_Z = \frac{\mu_{\max,Z} Z}{K_{S,Z} + Z} \tag{6}$$

$$r_P = \alpha \mu_L + \beta \qquad t \geq 24 \text{ h} \tag{7}$$

$$r_P = 0 \qquad t < 24 \text{ h}$$

Parameter Estimation

Parameters in the differential equations were estimated by a nonlinear least-squares routine (`lsqnonlin.m`) in the Optimization Toolbox (Coleman et al., 1999) of the MATLAB environment. The sum-squared error between the model predictions and the experimental data consisting of cell mass, xylose, lactose, and protein concentrations was minimized. Because the minimization of the sum of squares for a general nonlinear problem is quite dependent on the initial guesses for the parameters, various starting points were provided and the final parameter set that had the lowest sum-squared error was chosen. In addition, the initial conditions for the xylose and protein concentrations were assumed to be at an average value of 10.68 g/L and 0.75 g/L (obtained by averaging the initial values for these variables from the various experiments), respectively.

The estimated parameters are summarized in Table II.

Model Evaluation

The performance of this model is illustrated in Figure 2, which shows a comparison of the model predictions and experimental data for experiments carried out at initial lactose levels of approximately 1%, 3%, and 5% w/v. As this figure shows, the dynamics of lactose and xylose utilization are captured well, but the cell mass and protein production dynamics are not. Clearly, this very basic model does not have enough features to describe effectively the dynamic behavior of the system. In particular, the protein dynamics are not well described. This result provides the motivation for developing a more descriptive structured kinetic model.

Structured Kinetic Model

Model Development

The structured kinetic model incorporates a limited amount of structure in the cell mass component as depicted in Figure 3. As suggested by Velkovska et al. (1997) and Matsumura et al. (1981), cell mass is divided into three categories—primary mycelia, secondary mycelia, and spores. The governing differential equations describing this system are as follows:

$$\frac{dL}{dt} = -\mu_L X_p / Y_L \quad (8)$$

$$\frac{dZ}{dt} = -\mu_Z X_p / Y_Z \quad (9)$$

$$\frac{dX_p}{dt} = (\mu_L + \mu_Z - k_1 - k_{d1}) X_p \quad (10)$$

$$\frac{dX_s}{dt} = k_1 X_p - (k_2 + k_{d2}) X_s \quad (11)$$

$$\frac{dX_o}{dt} = k_2 X_s \quad (12)$$

$$\frac{dX}{dt} = \frac{dX_p}{dt} + \frac{dX_s}{dt} + \frac{dX_o}{dt} = (\mu_L + \mu_Z) X_p - k_{d1} X_p - k_{d2} X_s \quad (13)$$

$$\frac{dP}{dt} = r_P X_s \quad (14)$$

where L, Z, X, and P are the lactose, xylose, cell mass, and protein concentrations, respectively, and X_p, X_s, and X_o are the cell mass contributions from primary mycelia, secondary mycelia, and spores. The specific growth rates on lactose and xylose are μ_L and μ_Z, respectively; Y_L and Y_Z are the corresponding cell mass yields on lactose and xylose. k_1 and k_2 are constant rate terms for the conversion of primary mycelia to secondary mycelia and for the conversion of secondary mycelia to spores, respectively. k_{d1} and k_{d2} are endogenous death terms and r_P is the specific protein production rate. The postulated forms for the specific growth rates and production rate terms are the same as in the unstructured kinetic model.

Table IV. Normalization values used in the neural network model.

Variable name	Symbol	Normalization value
Time	t	120 h
Lactose concentration	L	60 g/L
Xylose concentration	Z	12 g/L
Cell density	X	28 g/L
Protein concentration	P	5 g/L

Parameter Estimation

A similar procedure as outlined in the section for the estimation of parameters in the unstructured model is followed. The estimated parameters are summarized in Table III.

Model Evaluation

The performance of this model is illustrated in Figure 4. As with the unstructured kinetic model, the dynamics of lactose and xylose utilization are captured fairly well. Although the protein dynamics track the data somewhat better than in the unstructured kinetic model, incorporation of the different cell growth stages into the model equations does not dramatically improve predictions about the dynamics of protein production. The assumption of a particular functional form for r_P in both the unstructured and structured models limits our ability to adequately simulate the dynamics of protein production.

Neural Network Parameter Function Model

Model Development

As presented by Tholudur and Ramirez (1996, 1999), the neural network parameter function methodology of modeling dynamic systems is a hybrid modeling technique that uses fundamental process knowledge in the form of conservation equations in combination with the function approximation capabilities of neural networks. Material balances on each major component being produced or consumed in the batch fermentations result in the following governing differential equations:

$$\frac{dL}{dt} = -r_L X \quad (15)$$

$$\frac{dZ}{dt} = -r_Z X \quad (16)$$

Table V. Neural network training details.

Function	Inputs	RMSE
r_L	t, L	3.1×10^{-3}
r_z	L, Z	1.1×10^{-3}
r_x	t, L,Z,X	1.1×10^{-4}
r_P	t, X, P	1.7×10^{-4}

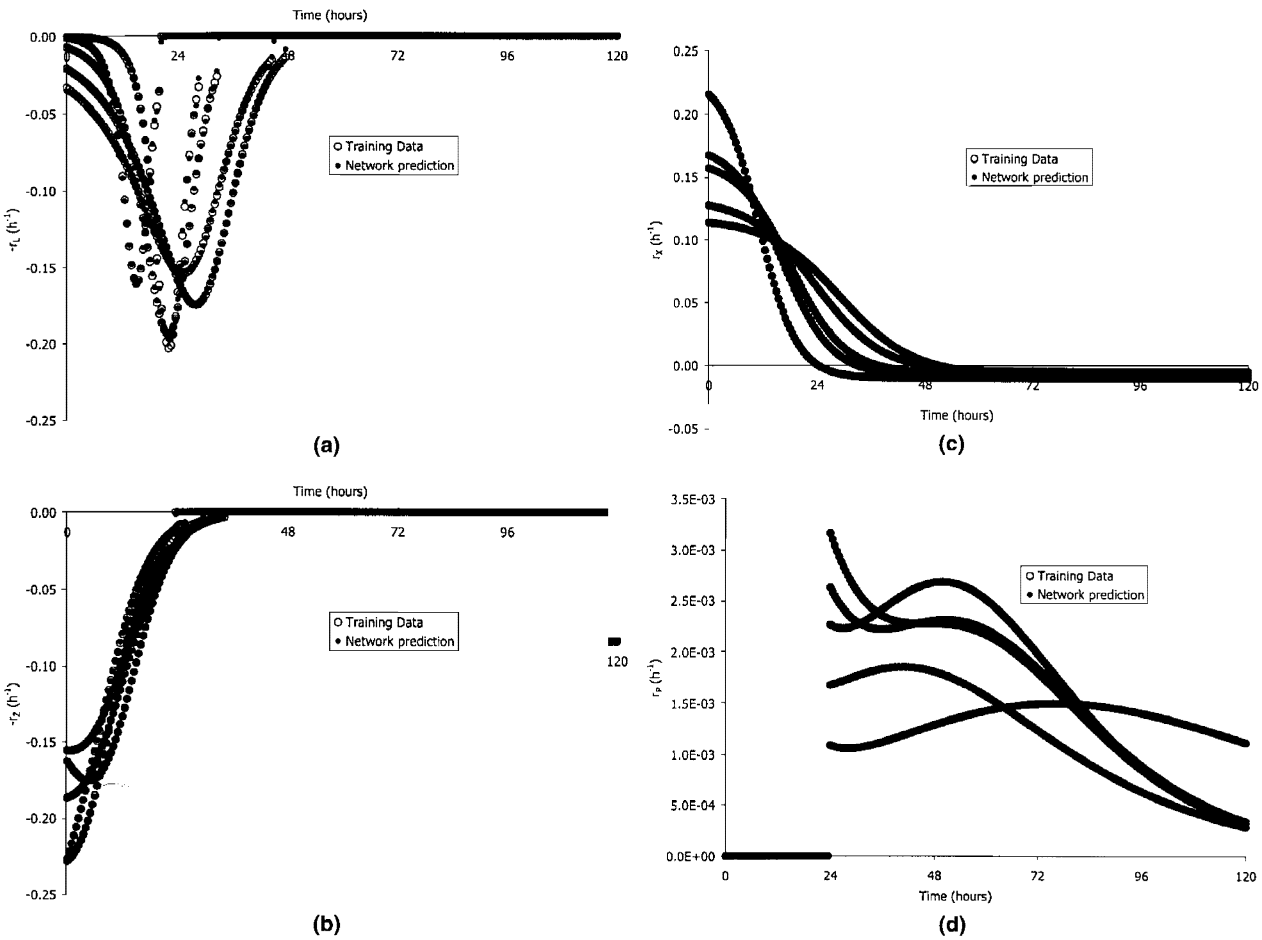


Figure 5. (a) Comparison of neural network predictions and actual training data for parameter function, r_L. (b) Comparison of neural network predictions and actual training data for parameter function, r_z. (c) Comparison of neural network predictions and actual training data for parameter function, r_x. (d) Comparison of neural network predictions and actual training data for parameter function, r_P.

$$\frac{dX}{dt} = r_X X \tag{17}$$

$$\frac{dP}{dt} = r_P X \tag{18}$$

where L, Z, X, and P are the lactose, xylose, cell mass, and protein concentrations, respectively. The modeling problem is one of estimating the parameter functions r_L, r_Z, r_X, and r_P, which are the specific lactose consumption rate, specific xylose consumption rate, specific cell growth rate, and specific protein production rate, respectively.

In the traditional kinetic modeling approaches, as presented in the earlier sections, a functional form is postulated for each function (e.g., Monod growth kinetics or a combination of primary and secondary metabolite formation), and the parameters in these functions are then evaluated. In the neural network parameter function modeling technique, the conservation equations are rewritten and solved for the unknown functions as follows:

$$-r_L = \frac{dL}{dt}/X \tag{19}$$

$$-r_Z = \frac{dZ}{dt}/X \tag{20}$$

$$r_X = \frac{dX}{dt}/X \tag{21}$$

$$r_P = \frac{dP}{dt}/X \tag{22}$$

Training data are then generated using these relationships. Specific algebraic input–output forms are assumed for each unknown function r_L, r_Z, r_X, and r_P. The methodology used to estimate the parameter functions is described in what follows.

Time Derivative and Parameter Function Evaluations

For the evaluation of the parameter functions, the time derivatives of the measured variables need to be estimated. To

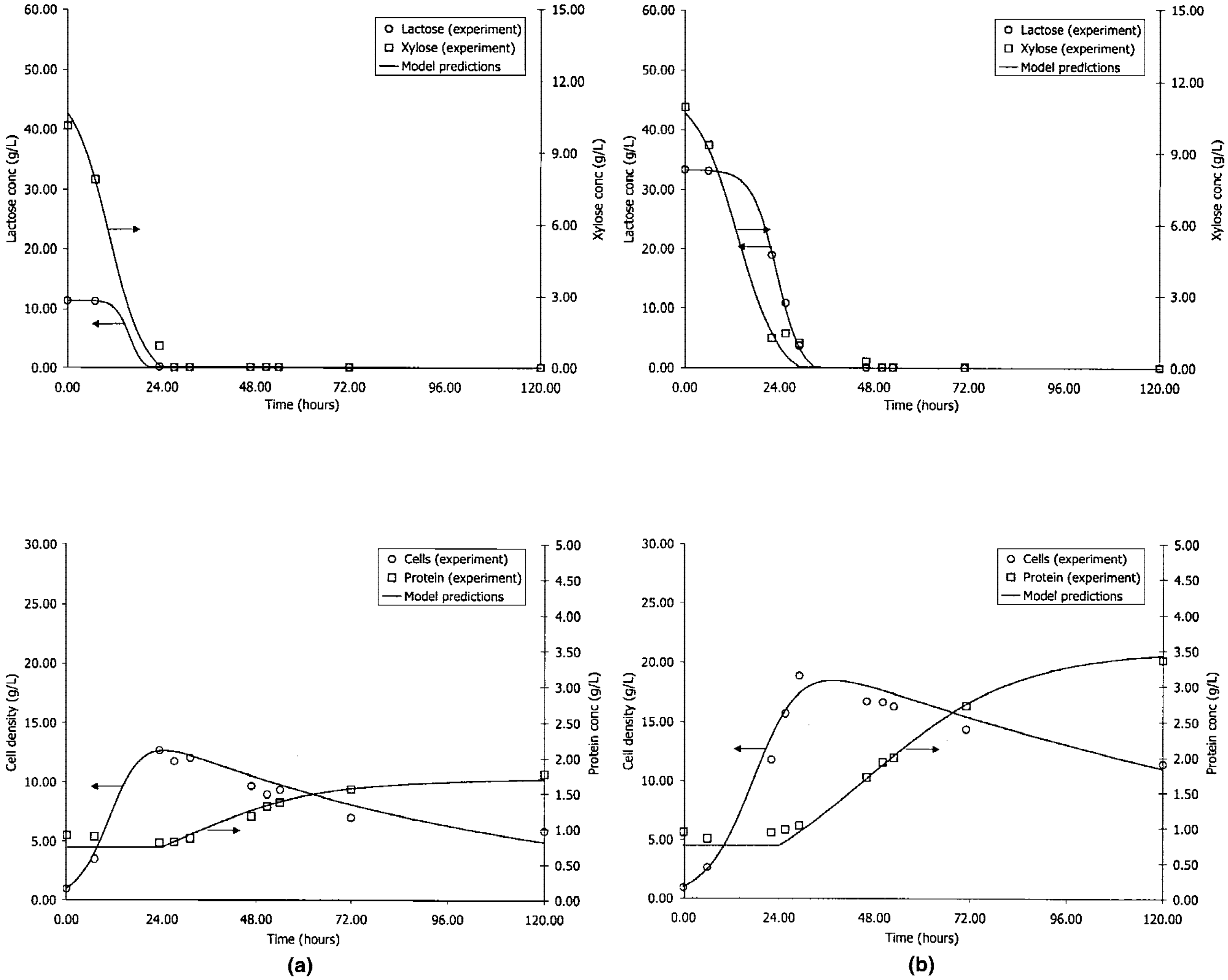


Figure 6. (a) Growth and production kinetics as predicted by the neural network-based model (initial xylose = 1% w/v, initial lactose = 1% w/v). (b) Growth and production kinetics as predicted by the neural network-based model (inital xylose = 1% w/v, initial lactose = 3% w/v). (c) Growth and production kinetics as predicted by the neural network-based model (initial xylose = 1% w/v, initial lactose = 5% w/v).

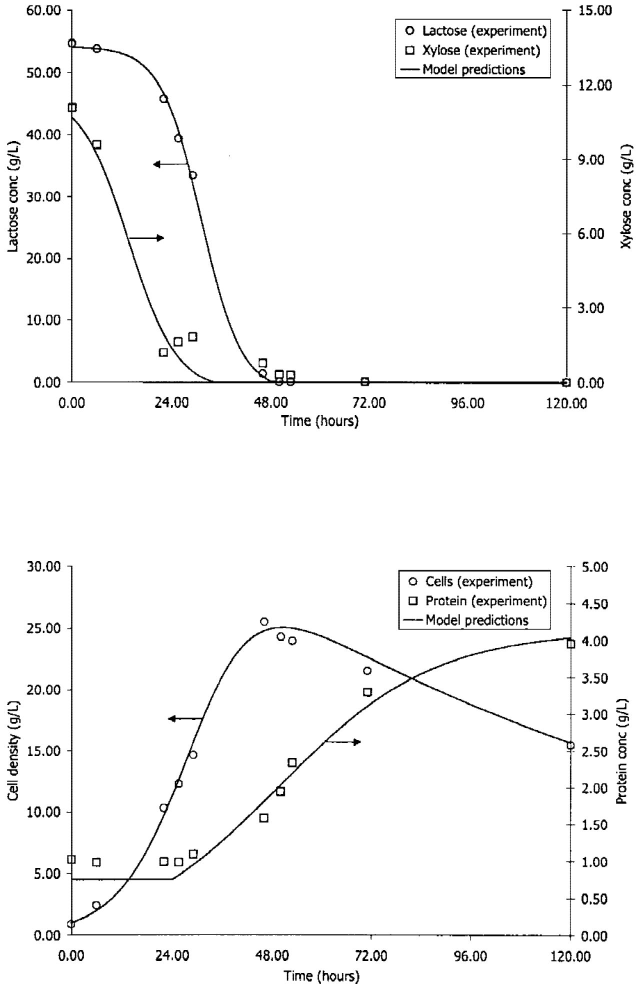
obtain these derivatives, we performed curve fits of the original experimental data and analytically differentiated the curve fits to obtain the derivatives. Protein production exhibits sigmoidal functional characteristics. Lactose and xylose concentrations exhibit monotonically decreasing sigmoidal behavior with respect to time. Thus, a logistic curve of the following form is fit for each of these variables to estimate the derivatives:

$$V(t) = \frac{c_1}{1 + e^{-c_2(t-c_3)}} \tag{23}$$

To better explain the protein dynamics, a lag term is also introduced (no protein production during the lag portion) to improve estimation of the derivatives. The cell mass concentration grows exponentially at first but fall once the sugars are consumed. The logistic equation presented above cannot describe these dynamics. Hence, a modified logistic function with one additional parameter is used:

$$V(t) = \frac{c_1 e^{-c_4 t}}{1 + e^{-c_2(t-c_3)}} \tag{24}$$

The curve fit, derivative estimation, and parameter function evaluation procedures were performed independently for each experimentally obtained data set. A sampling time of 0.5 h was used during this process—that is, the derivatives and parameter functions were estimated at 0.5-h intervals. This leads to a training data set that consists of 1205 input–output pairs for each parameter function. Such a short sampling time is necessary to justify the assumption of piecewise constant parameter functions between sampling points. In addition, the data were first normalized by dividing the variables with the appropriate normalization factor so that the magnitude of all the variables would be between 0 and 1. This helped in providing better fits of the logistic equations as well as in training the neural networks in the next stop. Table IV summarizes the normalization values that were used.



(c)

Figure 6. Continued.

Neural Network Training

After estimating the dynamic response data, the next step is to train the parameter function neural networks to mimic the training data. Because the parameter functions are static mappings (between the parameter functions and the measured variables), feedforward neural networks were chosen. A key issue in neural network training is the concept of generalization. A measure of the ability of the neural network to predict based on data it has not seen before is the generalization error (Geman et al., 1992; Moody, 1992). A lower generalization is always desirable. One method for improving generalization is to use a process known as regularization. The MATLAB Neural Network Toolbox (Demuth and Beale, 1998) provides a routine that trains neural networks using a combination of Bayesian regularization (MacKay, 1992) and Levenberg–Marquardt (Hagan and Menhaj, 1994) optimization. This routine (`trainbr.m`) was used to train the neural network parameter functions.

The choice of inputs for each network is another key issue. A correlation coefficient of all the data was estimated and used as a guide to choose the inputs for each network. Variables that were highly correlated with a particular parameter function were used as inputs for the corresponding parameter function neural network. Four neural networks were trained—one for each parameter function. All neural networks were trained with 20 hidden neurons (hyperbolic tangent activation function) and 1 output neuron (logistic activation function) for 1000 epochs. Table V summarizes the inputs used in the neural networks and the training performances achieved. A lower root-mean-squared error (RMSE) implies a better training performance. In addition, it is of interest to note the use of t as one of the inputs for three of the networks, the exception being r_Z. The inclusion of time has been observed earlier with the inducible bacteria systems (Tholudur and Ramirez, 1999).

The performance of the neural networks is also illustrated in Figure 5, which compares the network performance with the actual training data for each network. The figure shows that the parameter function neural networks capture the training data well.

Model Evaluation

Once the neural network parameter functions are trained, the next step is to examine the predictive capabilities of the complete model. This is summarized in Figure 6. As can be seen, in addition to capturing the dynamics of lactose and xylose consumption, the neural network-based model describes the dynamics of cell growth and protein production well. Comparison of Figures 2, 4, and 6 shows that the neural network-based model best captures the dynamics of the system among the models presented. In addition, the predictive capabilities of the neural network-based model on the remaining experimental data are shown in Figure 7. Quantification of the predictive capabilites of the various models is summarized in Table VI where the RMSE of the models for each of the state variables is presented. The consistently lower RMSE for the neural network model suggests that this model is more accurate than the kinetic models.

INTERPOLATED PARAMETER FUNCTIONS

The development of kinetic and neural network parameter function models for describing the dynamics of cell growth and cellulase enzyme production was presented in the previous section. The performance of these models on the experimental data was demonstrated. The optimization method used in the next section to obtain optimal operational conditions requires the model predictions at intermediate conditions as well (i.e., for initial lactose concentrations other than those for which experiments were conducted). In the case of the kinetic models, this is a trivial task because of the assumption of continuous parameter functions. However, in the case of the neural network parameter function models, the neural network might not be

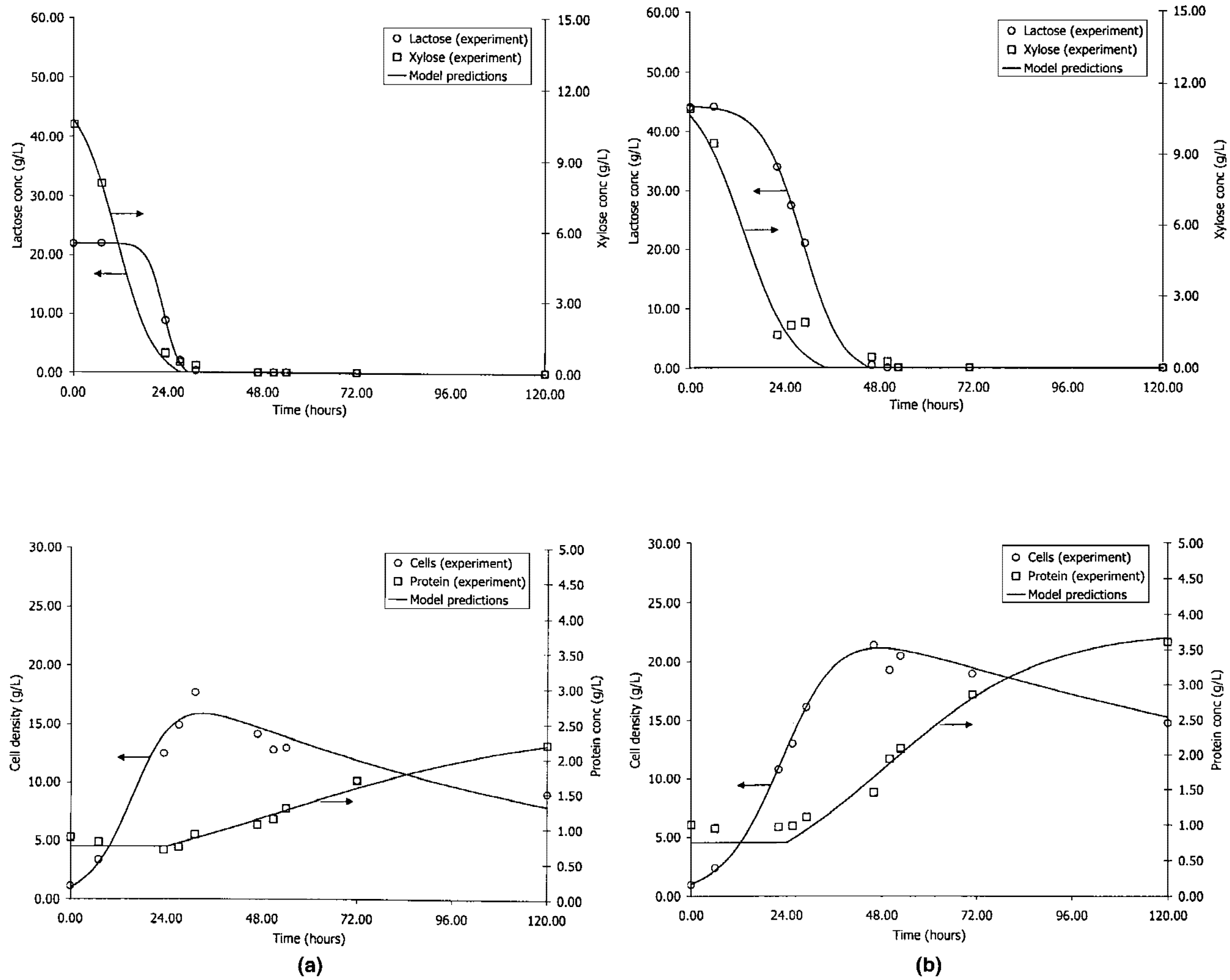


Figure 7. (a) Growth and production kinetics as predicted by the neural network-based model (initial xylose = 1% w/v, initial lactose = 2% w/v). (b) Growth and production kinetics as predicted by the neural network-based model (initial xylose = 1% w/v, initial lactose = 4% w/v).

able to predict reasonable values for some of these conditions.

The concept of interpolated parameter functions is used to obtain dynamic profiles of the states for conditions other than those for which experiments were conducted. For example, in this study, experiments were conducted at initial lactose concentrations of approximately 1%, 2%, 3%, 4%, and 5% w/v. The neural network parameter function model works extremely well for these initial conditions. To obtain the dynamic response for an initial lactose concentration of, say, 3.7% w/v, a weighted average of the parameter functions for the two nearest experimental points (3% and 4%) is used. This is different than a weighted average of the profiles. The interpolated parameter function method provides us with a means to use the neural network parameter functions with very few experiments. This methodology is summarized next.

Assume that N experiments are conducted by varying the initial conditions C_j, $j = 1, 2, 3, \ldots, N$. The parameter functions profiles, $F_j(t)$, are obtained for these conditions using the techniques just explained. To obtain the parameter

Table VI. Quantification of the predictive capabilities of the various models.

Model name	Root-mean-squared error				
	Lactose	Xylose	Cell Mass	Protein	Total
Unstructured kinetic	1.3887	0.6039	0.9518	0.2113	1.8011
Structured kinetic	1.3478	0.6129	0.9603	0.2047	1.7766
Neural network	0.4202	0.4560	0.9204	0.1346	1.1179

Table VII. Parameters used in the performance index expressions.

Parameter	Value
V	3 L
C_L	\$0.15/lb.
C_P	\$4/lb.
$Q_{\max}$	0.0615 g/(L · h)
$P_{\max}$	12 g

function profile, $F_I(t)$, for an initial condition, C_I, determine M ($1 \leq M \leq N-1$) such that $C_M \leq C_I \leq C_{M+1}$. Then, define $\kappa = C_I - C_M/C_{M+1} - C_M$. The interpolated function is defined as $F_I(t) = (1-\kappa)F_M(t) + \kappa F_{M+1}(t)$. Repeat this process for all the parameter functions. These interpolated parameter functions are then used in the governing differential equations, which are integrated to provide the dynamic profiles of the states for the intermediate initial condition.

The main advantage of this approach is the ability to obtain accurate interpolated information with a minimum of experimental data. For the neural networks to be able to interpolate directly would require a great deal more experimental training data sets.

MODEL OPTIMIZATION

One objective for developing predictive dynamic models of processes is to use these models for process optimization. Optimization requires identifying the critical process variables, formulating performance indices that reflect the interactions among these variables, and finally choosing a strategy to maximize or minimize the performance indices.

In this particular process, two performance indices can be readily formulated, involving lactose, protein, and time as the critical variables. The first performance index, J_1, balances the estimated protein value against the cost of lactose and can be written on a total protein basis as:

$$J_1 = P(t)V - \frac{C_L}{C_P} L(t_0)V \tag{25}$$

where $P(t)$ is the protein concentration at time t, $L(t_0)$ is the lactose concentration at time $t = 0$, V is the fermentation volume, and C_L and C_P are the cost of lactose and value of protein, respectively. Thus, this performance index provides a measure of the tradeoff between the value of the protein and the operating cost of using lactose as a carbon source and expresses the result in terms of net valuable protein (g).

Another important performance index is the average volumetric productivity of the system. Due to the absence of significant protein production during the first 24 h of the fermentation, the productivity performance index is defined as:

$$J_2 = \frac{P(t) - P(t = 24)}{t} \qquad t \geq 24 \tag{26}$$

This performance index has units of grams per liter per hour [g/(L·h)].

It is possible (and often very probable) that the optima resulting from either of these two performance indices would be different. Under such conditions, it might be necessary to arrive at a compromise. A weighted overall performance index of the following form can then be considered:

$$J = \gamma \frac{J_2}{Q_{\max}} + (1-\gamma) \frac{J_1}{P_{\max}} \tag{27}$$

where $Q_{\max}$ and $P_{\max}$ are normalization factors corresponding to maximal productivity and protein amount and γ is a parameter that expresses the relative weightings of the two performance indices.

For the purpose of illustration, the parameters in Table VII were used to calculate values of the three performance index expressions.

The results of the optimization of the performance indices, J_1 and J_2, using the three process models, are summarized in Figure 8. This figure clearly shows the existence of different and competing optima for the two performance indices J_1 and J_2. For example, the J_1 and J_2 contours for the unstructured kinetic model suggest that the optimum for maximizing J_1 is at about 5.5% w/v initial lactose and a final time of 120 h, whereas the optimum for maximizing J_2 is at about 5.5% w/v initial lactose and a final time of 78 h. Similarly, the contours for the structural kinetic model suggest an optimum for J_1 at about 5.5% w/v initial lactose level and a final time of 120 h, whereas the optimum for J_2 is at a similar lactose level but a reduced final time of about 83 h. And, finally, the contours obtained using the neural network model suggest an optimum of about 3.5% w/v initial lactose and a final time of 120 h for maximizing J_1, but an optimum of about 5.4% w/v initial lactose and a final time of about 78 h for maximizing J_2. Table VIII summarizes the optimal conditions obtained for each of these indices.

Comparison of the protein profiles predicted by the models shows that the unstructured and structured kinetic models exhibit significantly slower dynamics than seen in the experimental data. Hence, optimization of these dynamic performance indices using these models is not expected to provide accurate results. Considering that the dynamic behavior of the system is best captured by the neural network model, it is expected that the results of optimizing the neural network model be most accurate among the developed models. This is further demonstrated in Figure 9, which compares the performance indices predicted by the models with the experimental data. In Figure 9a, the values of the performance index, J_1, as predicted by the various models at a fixed final time of 120 h and as a function of the initial lactose concentration are shown. In Figure 9b, the values of the performance index, J_2, as predicted by the various models at a fixed initial lactose concentration of about 5.5% w/v, and as a function of time, are shown. The neural network-based model predicts performance index profiles that are closest to the experimental data. For the performance index predictions shown in Figure 9, the root-mean-squared errors are calculated and summarized in Table IX. Once again, the neural network-based model has the lowest RMSE in the optimal performance index predictions. As a result, it can be concluded that the opimization results from the neural network-based model are closest to the true optimum.

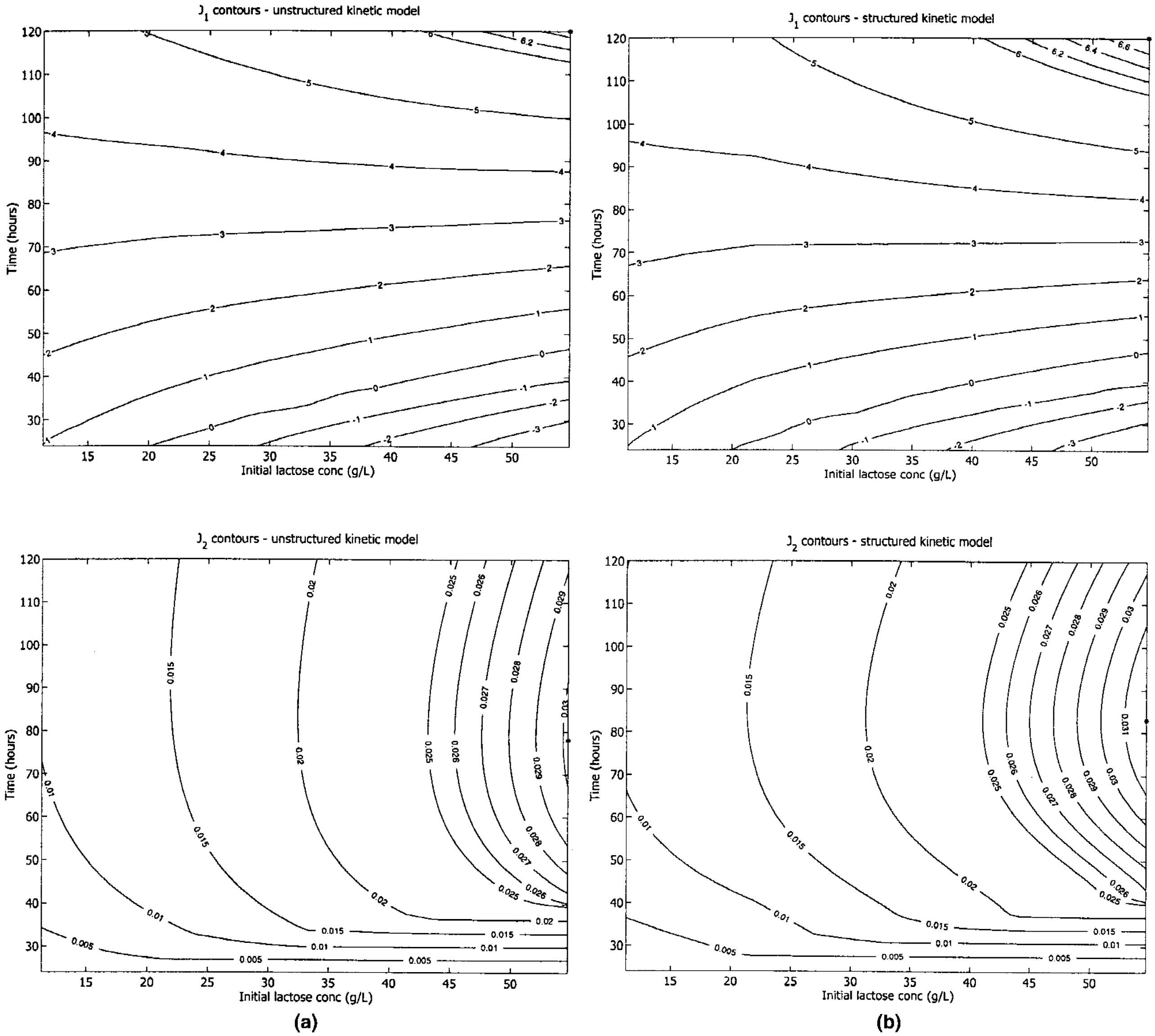


Figure 8. (a) Contour plots of the two performance indices using the unstructured kinetic model. (b) Contour plots of the two performance indices using the structured kinetic model. (c) Contour plots of the two performance indices using the neural network-based model.

CONCLUSIONS

The development and comparison of traditional kinetic modeling and neural network parameter function modeling as applied to a biotechnology process involving the production of cellulase proteins has been presented. Experimental data were used in the estimation of parameters in the various models. The models were developed using a similar amount of prior knowledge of the process (in terms of basic material balances). The predictive capabilities of the neural network-based model were superior to the unstructured and structured kinetic models developed. This is borne out by the consistently lower RMSE values in the model predictions for the process state variables. An approximately 33% lower RMSE in protein predictions and about 40% lower total RMSE was obtained with the neural network-based model as opposed to the kinetic models. Optimal conditions for the operation of the batch system were identified to demonstrate the practical applicability of modeling and optimizing system performance using the neural network parameter function modeling techniques. Once again, a comparison with the experimental data revealed that the neural network-based model predicted optimal conditions that were closer to the experimental observations than the other models. Using the neural network-based model, the RMSE in predicting the J_1 and J_2 preformance indices were about 67% and 40% lower, respectively, when compared with the kinetic models.

The neural network parameter function modeling scheme does suffer from the drawback that it does not provide any insight into the underlying mechanisms that describe the system. A similar observation can be made for the unstruc-

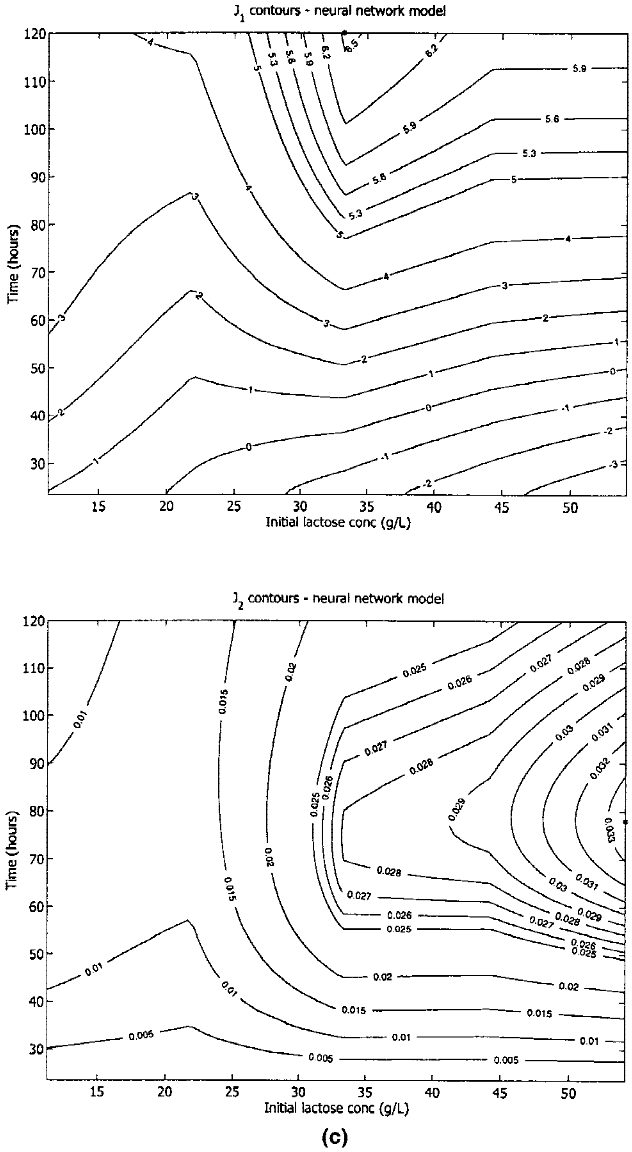


Figure 8. Continued.

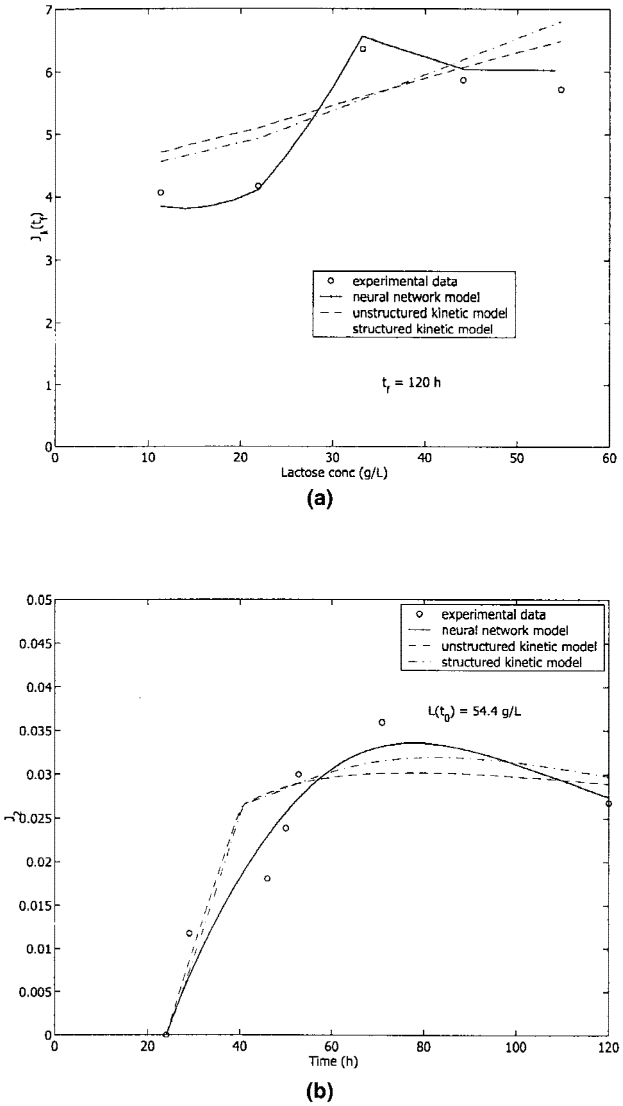


Figure 9. Comparison of optimal predictions of the models with experimental data.

tured kinetic model. Of the three models developed in this work, the structured kinetic model is probably the best in terms of describing the underlying phenomena. However, the neural network based method has the advantage of being able to be developed quite rapidly, thus enabling quicker optimization of the process.

A simplified system utilizing soluble sugars as the carbon source for the organism was used in this work. Currently, we are working on applying the modeling principles outlined in this work to a more complex insoluble substrate system.

Table VIII. Optimal conditions predicted by the different models.

Model name	Performance index	Maximum value	Optimal conditions $L(t_0)$ (g/L)	Optimal conditions t_f (h)
Unstructured kinetic	J_1	6.49 g	54.70	120.0
Structured kinetic	J_1	6.80 g	54.70	120.0
Neural network	J_1	6.56 g	33.25	120.0
Unstructured kinetic	J_2	0.0302 g/(L · h)	54.70	78.0
Structured kinetic	J_2	0.0319 g/(L · h)	54.70	83.0
Neural network	J_2	0.0336 g/(L · h)	54.08	78.0

Table IX. RMSE values for the performance index predictions in Figure 9.

Model name	Performance index	Prediction RMSE
Unstructured kinetic	J_1	0.70 g
Structured kinetic	J_1	0.74 g
Neural network	J_1	0.21 g
Unstructured kinetic	J_2	0.0049 g/(L · h)
Structured kinetic	J_2	0.0048 g/(L · h)
Neural network	J_2	0.0031 g/(L · h)

The authors thank Ali Mohagheghi, Tammy Hayward, Jenny Hamilton, and Mark Ruth (all at NREL) for their input, help, and fruitful discussions. We also thank Bill Adney (NREL) for guidance in quantifying cellulase protein levels while using the "desalting HPLC" technique described earlier in the work.

NOMENCLATURE

c_1, c_2, c_3, c_4	parameters in the logistic equation
C_j	experimental conditions
F_j	parameter function profile corresponding to the experimental condition C_j
J	Weighted overall performance index
J_1	relative protein amount (g)
J_2	average volumetric protein productivity [g/(L · h)]
k_1	conversion rate of primary mycelia to secondary mycelia (h^{-1})
k_2	conversion rate of secondary mycelia to spores (h^{-1})
k_d	endogenous death rate (h^{-1})
k_{d1}	primary mycelial cell lysis rate (h^{-1})
k_{d2}	secondary mycelial cell lysis rate (h^{-1})
$K_{I,L}$	inhibition constant for xylose on lactose (g/L)
$K_{S,L}$	saturation constant for lactose (g/L)
$K_{S,Z}$	saturation constant for xylose (g/L)
L	lactose concentration (g/L)
P	hydrolytic protein concentration (g/L)
P_{max}	maximum protein amount (g)
Q_{max}	maximum volumetric protein productivity [g/(L · h)]
$-r_L$	specific lactose consumption rate (h^{-1})
r_P	specific protein production rate (h^{-1})
r_X	specific cell growth rate (h^{-1})
$-r_Z$	specific xylose consumption rate (h^{-1})
V	fermentor volume (L)
X	fungal cell concentration (g/L)
X_o	spore concentration (g/L)
X_p	primary mycelial concentration (g/L)
X_s	secondary mycelial concentration (g/L)
Y_L	yield of fungal cells on lactose (g/g)
Y_Z	yield of fungals cells on xylose (g/g)
Z	xylose concentration (g/L)
α	constant in specific protein production rate
β	constant in specific protein production rate (h^{-1})
γ	performance index weighting parameter
κ	interpolating weighting parameter
μ_L	specific growth rate on lactose (h^{-1})
$\mu_{max,L}$	maximum specific growth rate on lactose (h^{-1})
$\mu_{max,Z}$	maximum specific growth rate on xylose (h^{-1})
μ_Z	specific growth rate on xylose (h^{-1})

References

Acebal C, Castillon MP, Estrada P, Mata I, Aguado J, Romero D. 1988. Production of cellulases by *Trichoderma reesei* QM9414 in batch and fed-batch cultures on wheat straw. Acta Biotechnol 8:487–494.

Adney WS, Mohagheghi A, Thomas SR, Himmel ME. 1995. Comparison of protein contents of cellulase preparations in a worldwide round-robin assay. In: Saddler JN, Penner M, editors. Enzymatic degradation of insoluble polysaccharides. ACS series 618. Washington, DC: ACS. p 256–271.

Bader J, Klingspohn U, Bellgardt KH, Schugerl K. 1993. Modelling and simulation of the growth and enzyme production of *Trichoderma reesei* Rut C30. J Biotechnol 29:121–135.

Chaudhuri BK, Sahai V. 1993. Production of cellulase using a mutant strain of *Trichoderma reesei* growing on lactose in batch culture. Appl Microbiol Technol 39:194–196.

Chaudhuri BK, Sahai V. 1994. Comparison of growth and maintenance parameters for cellulase biosynthesis by *Trichoderma reesei* C-5 with some published data. Enzyme Microbiol Technol 16:1079–1083.

Coleman T, Branch MA, Grace A. 1999. Optimization toolbox—user's guide, v 2. Natick, MA: The MathWorks.

Demuth H, Beale M. 1998. Neural network toolbox—user's guide, v 3. Natick, MA: The MathWorks.

Esterbauer H, Steiner W, Labudova I, Hermann A, Hyan M. 1991. Production of *Trichoderma* cellulase in laboratory and pilot scale. Biores Technol 36:51–65.

Geman S, Bienenstock E, Doursat R. 1992. Neural networks and the bias/variance dilemma. Neural Comput 4:1–58.

Hagan MT, Menhaj M. 1994. Training feedforward networks with the Marquardt algorithm. IEEE Trans Neural Networks 5:989–993.

Hendy NA, Wilke C, Blanch HW. 1984. Enhanced cellulase production in fed-batch culture of *Trichoderma reesei* C30. Enzyme Microbiol Technol 6:73–77.

Kadam KL. 1996. Cellulase production. In: Wyman CE, editors. Handbook on bioethanol: production and utilization. Washington, DC: Taylor & Francis. p 213–252.

Lejeune R, Baron GV. 1995. Effect of agitation on growth and enzyme production of *Trichoderma reesei* in batch fermentation. Appl Microbiol Technol 43:249–258.

Lejeune R, Baron GV. 1998. Modeling the exponential growth of filamentous fungi during batch cultivation. Biotechnol Bioeng 60: 169–179.

MacKay DJC. 1992. Bayesian interpolation. Neural Comput 4:415–447.

Matsumura M, Imanaka T, Yoshida T. 1981. Modeling of cephalosporin C production and its application to fed-batch culture. J Ferment Technol 59:115–123.

Mohagheghi A, Grohmann K, Wyman CE. 1988. Production of cellulase on mixtures of xylose and cellulose. Appl Biochem Biotechnol 17: 263–277.

Mohagheghi A, Grohmann K, Wyman CE. 1990. Production of cellulase on mixtures of xylose and cellulose in a fed-batch process. Biotechnol Bioeng 35:211–216.

Moody JE. 1992. The effective number of parameters: an analysis of generalization and regularization in nonlinear learning systems. Adv Neural Inform Proc Syst 4:847–854.

Nielsen J. 1992. Modelling the growth of filamentous fungi. Adv Biochem Eng/Biotechnol 46:189–223.

Nielsen J. 1996. Modelling the morphology of filamentous microorganisms. TIBTECH 14:438–443.

Nielsen J, Villadsen J. 1992. Modelling of microbial kinetics. Chem Eng Sci 47:4225–4270.

Philippidis GP. 1994. Cellulose production technology—evaluation of current status. In: Himmel ME, Baker JO, Overend RP, editors. Enzymatic conversion of biomass for fuels production. Washington, DC: ACS.

Rakshit SK, Sahai V. 1991. Optimal control strategy for the enhanced production of cellulase enzyme using the new mutant *Trichoderma reesi* E-12. Bioproc Eng 6:101–107.

Schafner DW, Toledo RT. 1991. Cellulase production by *Trichoderma reesei* when cultured on xylose-based media supplemented with sorbose. Biotechnol Bioeng 37:12–16.

Schafner DW, Toledo RT. 1992. Cellulose production in continuous culture by *Trichoderma reesei* on xylose-based media. Biotechnol Bioeng 39:865–869.

Schell DJ, Hinman ND, Wyman CE, Werdene PJ. 1990. Whole broth cellulase production for use in simultaneous saccharification and fermentation. Appl Biochem Biotechnol 24/25:287–297.

Szengyel Z, Zacchi G, Reczey K. 1997. Cellulase production based on hemicellulose hydrolysate from steam-pretreated willow. Appl Biochem Biotechnol 63–65:351–362.

Tangnu SK, Blanch HW, Charles R. 1981. Enhanced production of cellulase, hemicellulase and β-glucosidase by *Trichoderma reesei* (Rut C-30). Biotechnol Bioeng 23:1837–1849.

Tholudur A, Ramirez WF. 1996. Optimization of fed-batch bioreactors using neural network parameter function models. Biotechnol Prog 12: 302–308.

Tholudur A, Ramirez WF. 1998. Neural network modeling and optimization of foreign protein production by recombinant bacteria. AIChE J 45:1660–1670.

Velkovska S, Marten MR, Ollis DF. 1997. Kinetic model for batch cellulase production by *Trichoderma reesei* RUT C30. J Biotechnol 54: 83–94.